Update on the National Science Board Priorities and Recommendations on Scientific Data Management

January 12, 2012 • CENDI Meeting

José-Marie Griffiths, PhD, Vice President for Academic Affairs
Bryant University, Smithfield, Rhode Island

NSB

BRYANT

Data Policy Task Force

- Established at the February 3-4, 2010 National Science Board (NSB) meeting
- Charge: further defining the issues and outlining possible options to make the use of data more effective in meeting the National Science Foundation's (NSF) mission.

Data Policy Task Force Strategies

Monitor Impact

Monitor the impact of NSF updated implementation of the Data Management Plan requirement to inform a review of NSF policy.

Statement

Considering issues of data policy, Open Data movements, and related issues, the Task Force will then develop a "Statement of Principles."

Guidance

Provide guidance to subsequent Board efforts to develop specific actionable policy recommendations focused, initially, on NSF, but could potentially promulgate through other Federal agencies in a national and international context.


1.Openness, Transparency
2. Open Access Publishing Links
3. Stakeholder participation
4. Accommodate variation
5. Open Sharing and Active Management
6. Define roles, responsibilities and resourcing
7. Investigator rights and responsibilities
NSB Task Force on Data Policy Statement of Principles

1. Openness and transparency are critical to continued scientific and engineering progress and to building public trust in the nation's scientific enterprise.
   This applies to all materials necessary for verification, replication and interpretation of results and claims, associated with scientific and engineering research.

2. Open Data sharing is closely linked to Open Access publishing and they should be considered in concert.

NSB Task Force on Data Policy Statement of Principles

3. The nation's science and engineering enterprise consists of a broad array of stakeholders, all of which should participate in the development and adoption of policies and guidelines.

4. It is recognized that standards and norms vary considerably across scientific and engineering fields and such variation needs to be accommodated in the development and implementation of policies.

Task Force on Data Policy
Statement of Principles

NSB Task Force on Data Policy
Statement of Principles

5. Policies and guidelines are needed for open data sharing which in turn requires active data management.

6. All data and data management policies must include clear identification of roles, responsi-bilities and resourcing.


7. Investigator rights and responsibilities
7. The rights and responsibilities of investigators are recognized. Investigators should have the opportunity to analyze their data and publish their results within a reasonable time.
NSB Task Force on Data Policy
Statement of Principles


1.Openness, Transparency
2. Open Access Publishing Links
3. Stakeholder participation
4. Accommodate variation
5. Open Sharing and Active Management
6. Define roles, responsibilities and resourcing
7. Investigator rights and responsibilities
NSB Task Force on Data Policy Statement of Principles

NSB Expert Panel Discussion on Data Policies

- March 28-29, 2011
- Arlington, VA
- Participants included:
  - Over 30 experts/research administrators
  - 7 NSB members
  - 4 NSF Directors/Staff

Standards and Interoperability Enable Data-intensive Science

- Need new norms and practices
- Data producers, software & tool developers, data curators get credit for their work

- To enable sharing & interoperability across disciplines and internationally

- To enable tracking of provenance
- Ensure data integrity (see next section)
- Facilitate citation & attribution

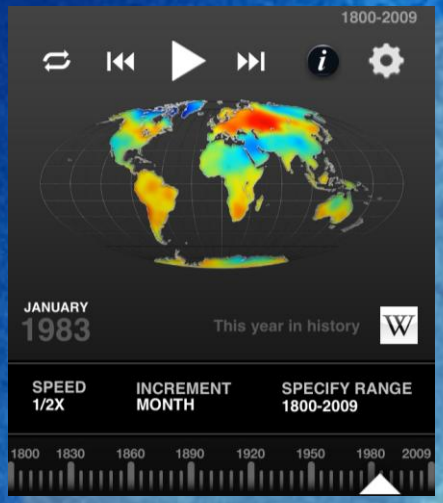
Data Sharing is an Identified Priority

- Must balance privacy concerns and data access for sharing and re-use.
- Acknowledge disciplinary cultures while establishing a culture of sharing across all research communities.
- Must promote & reward exemplary data management projects & plans.
- Data availability must be timely – issues of embargoes and restricted use durations.

Recognize and Support Computational and Data-intensive Science as a Discipline

- Recognize & reward computational & data scientists & curators: funding, tenure, etc.
- Support training in computational science
- Reward international collaborations to develop cyberinfrastructure, data stewardship, interoperability, international sharing
- New funding/economic models to support processing, storing, archiving, maintaining data sets
- Need to define who is responsible for what – funding agencies/publishers versus research communities

Integrity Concerns for Research Institutions



- What to share - raw, processed, analyzed datasets, instruments, calibration and environmental records, analytical tools, etc.
- Processes for and costs of long-term curation of data

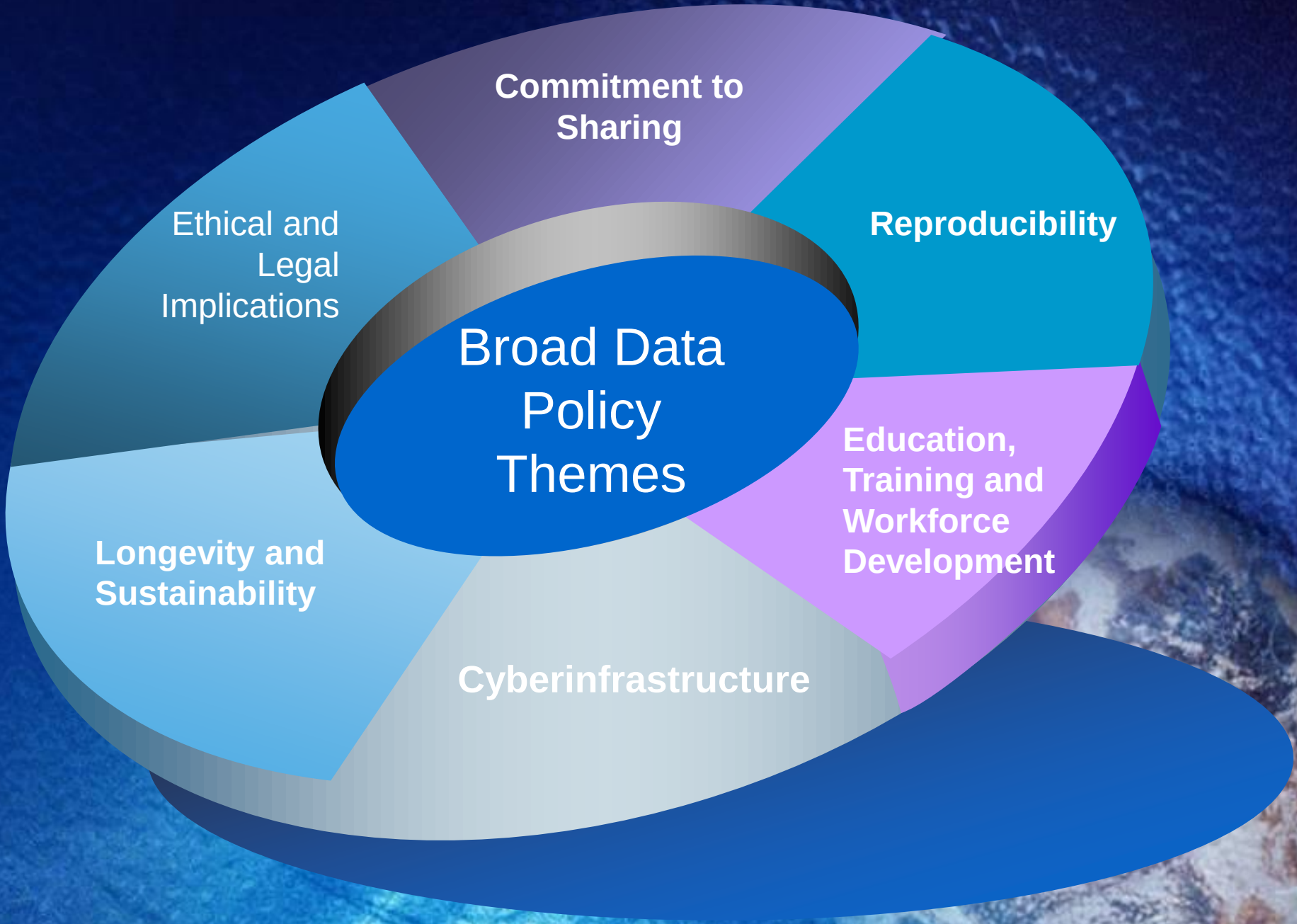
Storage, Preservation, and Curation of Data are Critical to Data Sharing and Management (data stewardship)

- Funding agencies must commit to ongoing financial support for repositories (no “orphans”)
- Standardized curatorial mechanisms
- Strategic partnerships between stakeholder communities and data repositories, supported by funders
- Define roles of different types of digital repositories
- Possibly independent auditing of data repositories to ensure data quality, access, interoperability

Cyberinfrastructure is Necessary to Support Data-intensive Science

- Geographic distribution of research teams, computing resources and datasets requires robust cyberinfrastructure
- Must include shared applications for analysis, visualization and simulation
- Standardization for interoperability & accessibility

- Need capital investment in cyberinfrastructure
- Need to define appropriate ratio of infrastructure to research funding

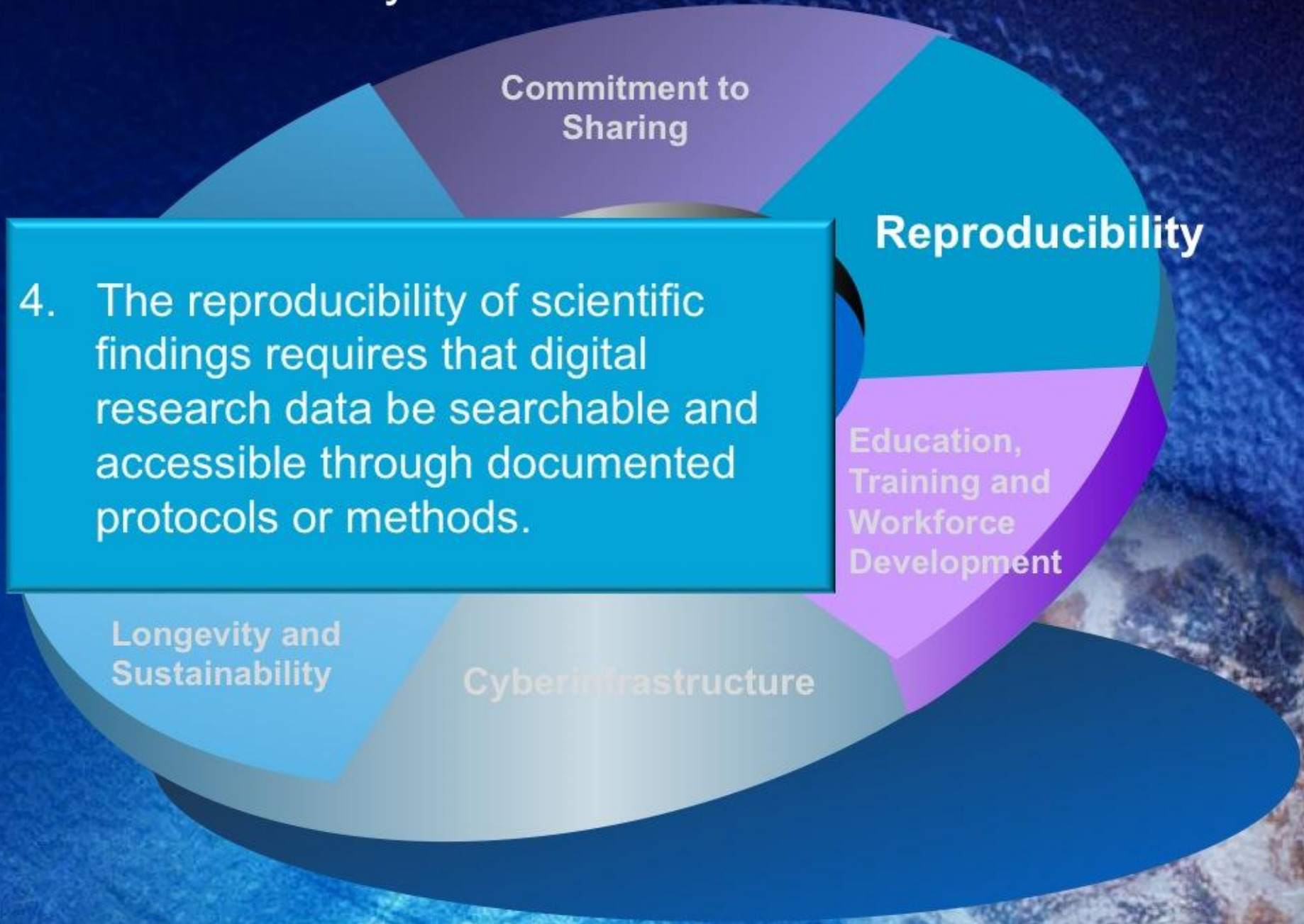

Commitment to Sharing
Reproducibility
Ethical and Legal Implications
Broad Data Policy Themes
Education, Training and Workforce Development
Longevity and Sustainability
Cyberinfrastructure

Broad Data Policy Themes

Commitment to Sharing

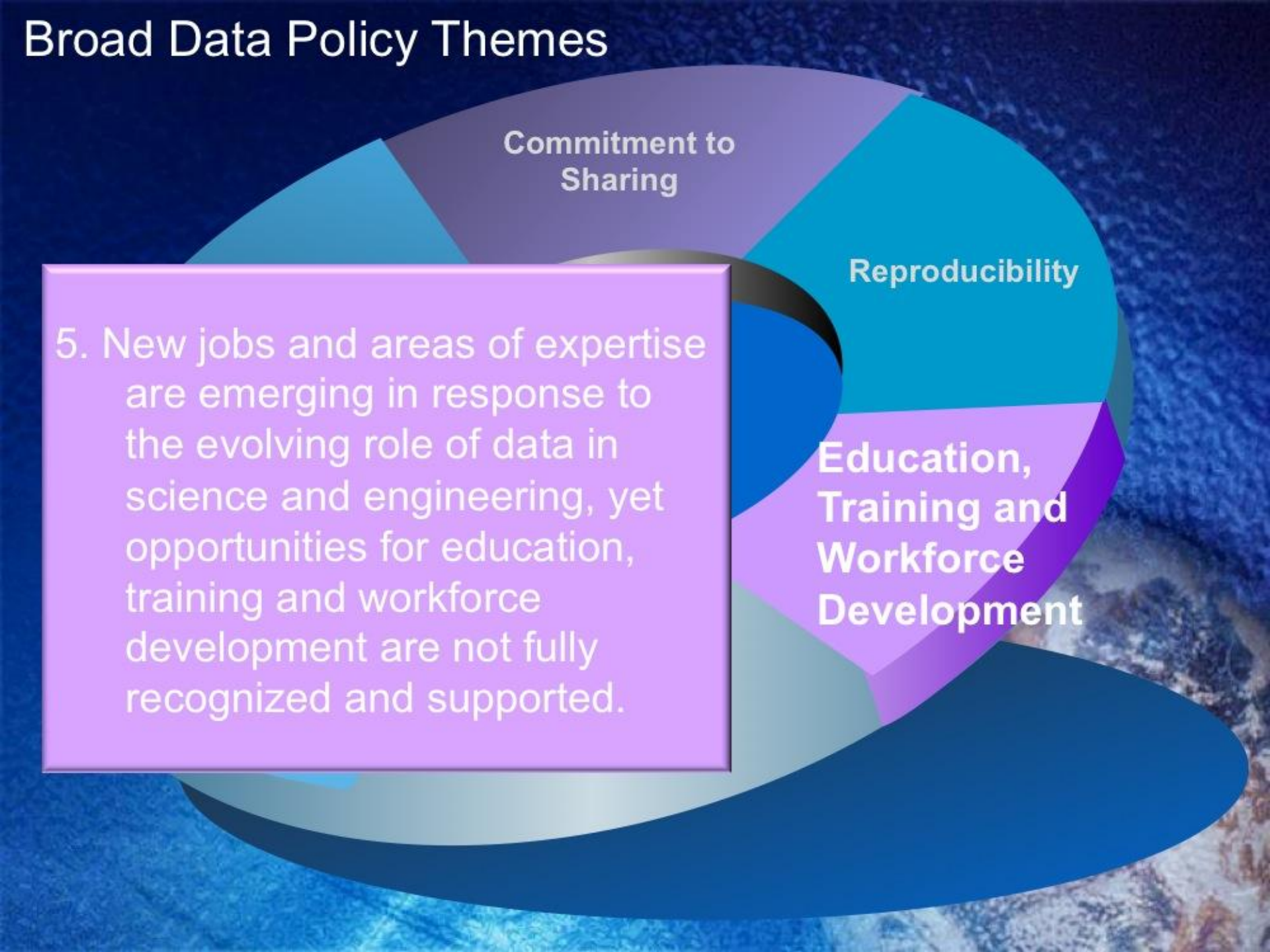
1. Broad stakeholder involvement and commitment to clear and realistic goals and measures of progress are necessary to ensure sustainable data sharing and management
2. As data collections expand in scale, scope and complexity, successful data sharing and management require a change in research and institutional cultures
3. Data sharing requires the coordination of goals and efforts through international collaborations and activities.

Broad Data Policy Themes



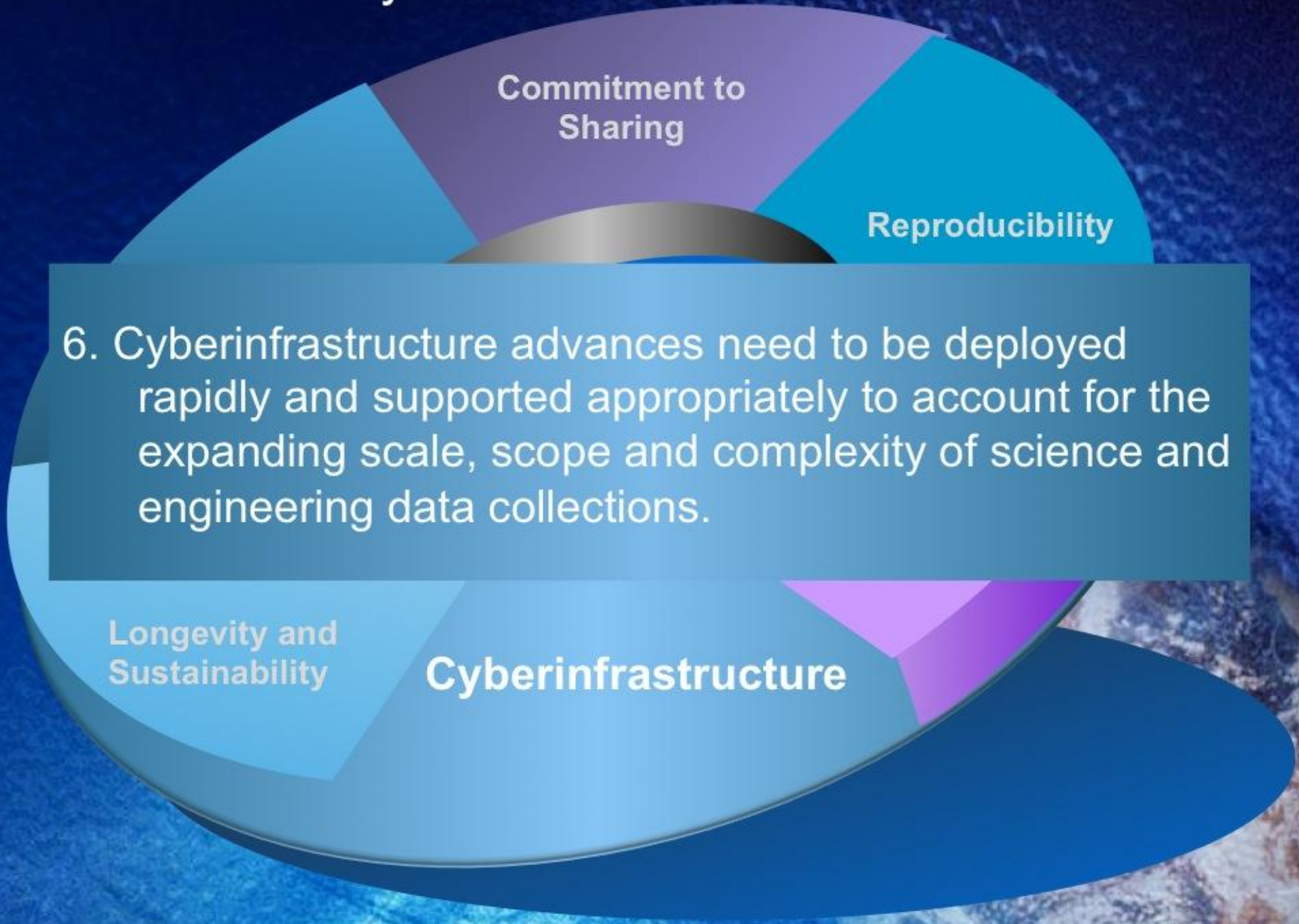
4. The reproducibility of scientific findings requires that digital research data be searchable and accessible through documented protocols or methods.

Broad Data Policy Themes



5. New jobs and areas of expertise are emerging in response to the evolving role of data in science and engineering, yet opportunities for education, training and workforce development are not fully recognized and supported.

Broad Data Policy Themes



6. Cyberinfrastructure advances need to be deployed rapidly and supported appropriately to account for the expanding scale, scope and complexity of science and engineering data collections.

Broad Data Policy Themes

7. Data stewardship is critical to the longevity and sustainability of data sharing and management throughout the data lifecycle, but it is unclear where the responsibilities for the effort lie.
8. Data stewardship must allow for broad and timely access to data.
9. Long-lived data require long-term business models that ensure data stewardship.

Longevity and Sustainability

Cyberinfrastructure

Broad Data Policy Themes

Areas of Recommendation

- Commitment to research data sharing
- Reproducibility
- Education, training and workforce development
- Longevity and sustainability

NSB Recommendations on Scientific Data Management


5.
Develop &
Implement Models
& Infrastructure
4.
Stakeholder
Models
Development
1. Leadership
Provide leadership to Federal agencies and other national and international stakeholders in the development and implementation of digital research data policies, including the promotion of individual scientific communities to establish data sharing and management practices that align with NSF data policies.
NSF
National
Science
Board
Recommendations on
Scientific Data
Management
1.
Leadership

2. Comprehensive Archiving

Consistent with the digital research data generated in research projects, require grantees to make both the data and the methods and techniques used in the creation and analysis of the data accessible for the purposes of building upon or verifying figures, tables, findings, and conclusions in peer-reviewed publications.

Similar requirements are appropriate when data are requested for the purpose of extending the scientific conclusions through further research. Data should be shared using persistent electronic identifiers, which enable automatic attribution of authors and award funding.

2. Comprehensive Archiving

1. Leadership

National Science Board Recommendations on Scientific Data Management

3. Computational Professionals Support

Continue to expand the support of computational and data-enabled science and engineering researchers and cyberinfrastructure professionals to take advantage of shared, accessible data and to forward emerging science.


4. Stakeholder Models Development
Convene a panel of stakeholders to explore and develop a range of viable long-term business models and issues related to maintaining digital data and provide a key set of recommendations for action.
5. Develop & Implement Models & Infrastructure
4. Stakeholder Models Development
2. Comprehensive Archiving
1. Leadership
NSF
National Science Board Recommendations on Scientific Data Management

5. Develop & Implement Models & Infrastructure

Further the expansion of sustainable data management, including preservation and curation of pre-existing and newly generated long-lived data, by encouraging development and implementation of data sharing infrastructure and long-term business models that encompass the range of research communities, research institutions, and research grants, as outlined in recommendations of the panel formed to explore these issues in Recommendation 4.

5.
Develop &
Implement Models
& Infrastructure

1.
Leadership

National Science Board Recommendations on Scientific Data Management


5.
Develop &
Implement Models
& Infrastructure
4.
Stakeholder
Models
Development
3.
Computational
Professionals
Support
2.
Comprehensive
Archiving
1.
Leadership
National
Science
Board
Recommendations on
Scientific Data
Management

José-Marie Griffiths, Ph.D.

- Vice President for Academic Affairs
  Bryant University
  1150 Douglas Pike
  Smithfield, RI 02917
  (401) 232-6060

  jmgriff@bryant.edu
  josemarie@gmail.com